AD-A049 784 STANFORD UNIV CALIF DEPT OF STATISTICS F/G 12/1
ESTIMATION FOLLOWING SEQUENTIAL TESTS.(U)
OCT 77 D SIEGMUND N00014-77-C-0306
UNCLASSIFIED TR-2 NL

| OF |

AD A049784

END
DATE
FILMED
3-78
DDC

12

ESTIMATION FOLLOWING SEQUENTIAL TESTS

BY

D. SIEGMUND

See back page for 1473

TECHNICAL REPORT NO. 2
OCTOBER 1, 1977

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ESTIMATION FOLLOWING SEQUENTIAL TESTS

by

D. Siegmund*

Technical Report No. 2

October 1, 1977

Prepared under Contract
N00014-77-C-0306 (NR-042-373)
for the Office of Naval Research
D. Siegmund, Project Director

*Research also partially supported by
PHS Contract 2 R01 GM16895 08 at
Columbia University.

Department of Statistics
Stanford University
Stanford, California

Abstract

Two methods for obtaining sequential test based confidence intervals for the mean of a normal distribution are proposed and compared. One method involves reducing the bias introduced into the sample mean by optional stopping. The other defines the confidence interval directly in terms of the stopping time of the test.

AMS Subject Classification. Primary 62L10

Key Words and Phrases: Sequential test, confidence interval, stopping rule.

ESTIMATION FOLLOWING SEQUENTIAL TESTS

1. Introduction

In decision oriented hypothesis testing sequential analysis offers an advantage over fixed sample size procedures. In cases which are not close to the borderline between hypotheses, one can make a decision on the basis of a relatively small amount of data and reduce the cost of experimentation. It is commonly supposed that sequential tests are not so useful when one wants to supplement his test results with an estimate of some parameter. The basis for this belief is the observation that sequential tests are designed to terminate sampling as soon as there is sufficient reason to reject one hypothesis, and it frequently occurs that insufficient data is then available for accurate estimation. A second reason for the paucity of studies of estimation following sequential tests is the mathematical difficulty associated with computing the sampling distributions of such estimates.

The primary purpose of this paper is to suggest point and confidence interval estimates for use in conjunction with the class of sequential tests recommended by Armitage (1975) for clinical trials. Two approaches are considered. One begins with the customary fixed sample size estimator which is then modified to reduce the bias created by random termination of the experiment. The other and more novel approach attempts to quantify one's intuitive feeling that the sooner a sequential test terminates the more evidence it provides for a large departure from the boundary between hypotheses. To facilitate understanding the proposed procedures most of the paper is restricted to estimating the mean of a normal population with known variance. The more practical and difficult cases of a normal mean with unknown variance and Bernoulli mean p are discussed briefly.

In clinical trials the desirability of sequential tests arises to a large degree from ethical considerations. For example, if one is comparing two treatments in a matched pairs experiment, where in each pair one person is assigned at random to each treatment, and if one treatment is considerably superior, the experiment should be terminated as soon as possible so that only a small number of patients are subjected to the inferior treatment. To be more specific, suppose that for $n = 1,2,\ldots$ the difference in response in the $n^{\underline{th}}$ pair, x_n, is normally distributed with unknown mean μ and known variance σ^2 independently of the differences in the other pairs. Let $s_n = x_1 + \ldots + x_n$, and for given $b > 0$ define the stopping rule

$$T = \text{first } n \geq 1 \text{ such that } |s_n| \geq \sigma b\sqrt{n} \quad . \tag{1}$$

Let m be a positive integer. Consider the sequential test of H_0: $\mu = 0$ against H_1: $\mu \neq 0$ which terminates sampling at $T' = \min(T,m)$ and rejects H_0 if and only if $T \leq m$.

Let $\theta = \mu/\sigma$. The distribution of T and hence the power function and expected sample size of these tests depend on the parameters (μ,σ) only through the value of θ. The tests have been studied numerically by McPherson and Armitage (1971)--see also Armitage (1975)--whose tables allow one to choose the design parameters b and m to attain a given significance level $\alpha = P_0\{T \leq m\}$ and type II error probability $\beta = P_\theta\{T > m\}$ at a given value $\theta \neq 0$. They have been studied asymptotically for large b and m by Siegmund (1977), whose approximations compare favorably with the numerical results of McPherson and Armitage, and whose methods form the basis for this paper. Technical aspects of the asymptotic analysis use ideas which go back to Anscombe (1953) and have recently also been studied by Woodroofe (1976b) in a context related to the present one.

The following heuristic considerations suggest that an experiment terminated by the stopping rule T may be useful in some estimation problems, and hence the goals of estimation and testing may not be in conflict to the degree commonly supposed.

The estimator $\bar{x}_n = s_n/n$ of μ based on $x_1,\ldots,x_n$ has variance σ^2/n. To estimate μ (assumed $\neq 0$) with a prescribed <u>proportional</u> accuracy, one would like to choose n so that $n^{-\frac{1}{2}}\sigma \leq \tilde{b}|\mu|$, where $\tilde{b}$ is a measure of the accuracy desired. This is, of course, impossible because the inequality defining n depends on the unknown value of $|\mu|$; but if one is free to choose his sample size sequentially, it suggests estimating $|\mu|$ by $|\bar{x}_n|$ and stopping with the first $n \geq 1$ such that $n^{-\frac{1}{2}}\sigma \leq \tilde{b}|\bar{x}_n|$. This stopping rule is of the form (1) with $\tilde{b} = b^{-1}$. Truncating the experiment at the $m^{\underline{th}}$ observation has the interpretation of accepting an estimate of prescribed absolute accuracy when $|\mu|$ appears to be so small that the prescribed proportional accuracy requirement would necessitate an extremely large sample size.

In Sections 2 and 3, where σ is assumed known, it will be convenient to take $\sigma = 1$, so $\theta = \mu/\sigma = \mu$.

2. Confidence Intervals

This section attempts to quantify by means of confidence intervals one's intuitive feeling that small values of the stopping rule T are evidence in favor of large values of $|\theta|$. It should be noted that with minor modifications the conceptual framework of the first part of this section may be adapted to various stopping rules and parent distributions. However, the asymptotic calculations that come later rely somewhat more heavily on the definition (1) and the assumption of normality. See also Section 4.

To determine confidence intervals for θ based on observing T' and $s_{T'}$ (here as before $T' = \min(T,m)$), it is useful to define several auxiliary functions. For $0 < \gamma \leq \frac{1}{2}$, $n = 1,2,\ldots$, and $|\xi| < b\sqrt{m}$, let

$$\underline{\theta}_1(n) = \inf\{\theta: P_\theta\{T \leq n,\ s_T > 0\} \geq \gamma\} \quad , \tag{2}$$

$$\underline{\theta}_2(\xi) = \inf\{\theta: P_\theta\{T \leq m,\ s_T > 0\} + P_\theta\{T > m,\ s_m \geq \xi\} \geq \gamma\} \quad , \tag{3}$$

$$\overline{\theta}_1(n) = \sup\{\theta: P_\theta\{T \geq n\} + P_\theta\{T < n,\ s_T < 0\} \geq \gamma\} \quad , \tag{4}$$

and

$$\overline{\theta}_2(\xi) = -\,\underline{\theta}_2(-\xi) \quad . \tag{5}$$

It is easy to see from continuity and monotonicity considerations that these functions are well defined and that $\underline{\theta}_i \leq \overline{\theta}_i$ $(i = 1,2)$. It is convenient to assume that $\underline{\theta}_1(m) \geq 0$, which will be the case provided that $2\gamma \geq P_0\{T \leq m\}$, the significance level of the test defined in Section 1. Let

$$\underline{\theta} = \underline{\theta}_1(T) I_{\{T \leq m, s_T > 0\}} + \underline{\theta}_2(s_m) I_{\{T > m\}} - \overline{\theta}_1(T) I_{\{T \leq m, s_T < 0\}} \tag{6}$$

and

$$\overline{\theta} = \overline{\theta}_1(T) I_{\{T \le m, s_T > 0\}} + \overline{\theta}_2(s_m) I_{\{T > m\}} - \underline{\theta}_1(T) I_{\{T \le m, s_T < 0\}} \quad . \tag{7}$$

Then $\underline{\theta} \le \overline{\theta}$ and the interval $[\underline{\theta}, \overline{\theta}]$ is a $100(1-2\gamma)\%$ confidence interval for θ. The proof of this last assertion is rather tedious, since it involves separate consideration of several cases. It is deferred to the Appendix.

Exact computation of the probabilities involved in (2)-(4) is a rather complicated numerical problem. A fairly easy asymptotic analysis yields what appear to be adequate approximations for a wide range of parameter values. Most of the rest of this section is devoted to describing this asymptotic analysis and to discussing an example.

In the approximate calculations that follow it is assumed that $b \to \infty$ and $m \to \infty$ in such a way that for some $\theta_1 > 0$

$$b = \theta_1 \sqrt{m} \quad . \tag{8}$$

It follows from (1) and the strong law of large numbers that as $b \to \infty$, for all $\theta > 0$

$$b\sqrt{T} \sim s_T \sim \theta T \quad \text{and hence} \quad T \sim (b/\theta)^2 \quad \text{a.s.} \tag{9}$$

Thus by (8)

$$P_\theta\{T \le m\} \to \begin{matrix} 0 & \text{for} & |\theta| < \theta_1 \\ 1 & \text{for} & |\theta| > \theta_1 \end{matrix} \quad , \tag{10}$$

so θ_1 defines asymptotically those values of θ for which the experiment is terminated before the m^{th} observation.

Let

$$T_+ = \inf\{n: s_n \ge b\sqrt{n}\} \quad . \tag{11}$$

The basic theoretical result of this paper is summarized in

Proposition 1. Under condition (8), for all θ and $0 < x < \infty$, as $b \to \infty$

$$P_\theta\{T_+ \le m, b\sqrt{m}-x \le s_m < b\sqrt{m}\} \sim m^{-\frac{1}{2}}\phi(\sqrt{m}(\theta-\theta_1))\int_0^x P_{-\frac{1}{2}\theta_1}\{\tau(y)<\infty\}\exp\{(\theta_1-\theta)y\}dy, \quad (12)$$

where $\tau(x) = \inf\{n: s_n > x\}$ and ϕ denotes the standard normal density function. For $\theta > 0, \ne \theta_1$ (12) also holds when $x = \infty$.

A heuristic argument in support of Proposition 1 is implicit in Section 3 of Siegmund (1977). A proof using a somewhat different approach is contained in the Appendix. Several of the important ideas go back to Anscombe (1953).

Of course, for use in calculating probabilities it is still necessary to evaluate the integral on the right hand side of (12). For $x = \infty$ and $\theta > \theta_1$ Siegmund (1977) observed that random walk theory allows one to rewrite this integral in a form suited for easy machine computation; and, in fact, the same formula is valid for $x = \infty$ and $0 < \theta < \theta_1$ also.

By standard reasoning for $\theta > 0$ as $x \to \infty$ $P_{-\theta}\{\tau(x) < \infty\} \sim \exp(-2\theta x)(P_{-\theta}\{\tau(0) = \infty\})^2/2\theta^2$ (e.g., Siegmund, 1975). Random walk theory yields a computationally convenient evaluation of $P_{-\theta}\{\tau(0) = \infty\}$ (Feller, 1966, p. 395) leading to the asymptotic relation

$$P_{-\frac{1}{2}\theta_1}\{\tau(x) < \infty\} \sim \nu(\theta_1)\exp(-\theta_1 x) \qquad (x \to \infty) \quad , \qquad (13)$$

where in terms of the standard normal distribution function Φ

$$\nu(\theta) = 2\exp\{-2\Sigma_1^\infty n^{-1}\Phi(-\tfrac{1}{2}\theta n^{\frac{1}{2}})\}/\theta^2 \quad . \qquad (14)$$

Table 1 gives selected values of ν.

TABLE 1

$\nu(\theta)$

θ	$\nu(\theta)$
.10	.9435
.15	.9164
.20	.8901
.25	.8645
.30	.8397
.35	.8157
.40	.7923
.45	.7696
.50	.7476
.55	.7263
.60	.7055
.65	.6854
.70	.6659
.75	.6470
.80	.6286
.85	.6107
.90	.5934
.95	.5767
1.00	.5604

Numerical evidence, some of which is mentioned in Siegmund (1975), suggests that the convergence indicated by (13) is very fast, and hence for $0 < x < \infty$ a reasonable approximation to the integral on the right in (12) is obtained by first writing $\int_0^x = \int_0^\infty - \int_x^\infty$ and then pretending that (13) is an equality to evaluate $\int_x^\infty$. For $\theta \leq 0$ and fairly small values of x it is probably reasonably accurate to substitute the right hand side of (13) directly into the integrand on the right in (12). Actually, even for $\theta > 0$ and $x = \infty$, when exact evaluation of the integral on the right hand side of (12) is possible, using (13) and the tabled values of (14) to obtain an approximate evaluation of (12) seems to be accurate enough for most practical purposes.

With the aid of Proposition 1 it is possible to give good approximations to the probabilities appearing in (2)-(4) for most parameter values of interest. For example, for $0 < \theta < \theta_1$

$$P_\theta\{T \leq m,\ s_T > 0\} \sim P_\theta\{T_+ \leq m\} = P_\theta\{s_m \geq b\sqrt{m}\} + P_\theta\{T_+ \leq m,\ s_m < b\sqrt{m}\} \quad ; \quad (15)$$

for $\theta > 0$ and $x > 0$

$$P_\theta\{T \leq m, s_T > 0\} + P_\theta\{T > m, s_m > b\sqrt{m}-x\} \cong P_\theta\{s_m \geq b\sqrt{m}-x\} + P_\theta\{T_+ < m, s_m < b\sqrt{m}-x\} \quad ; \quad (16)$$

and for $\theta > \theta_1$

$$P_\theta\{T > m\} + P_\theta\{T \leq m, s_T < 0\} \sim P_\theta\{T_+ > m\} = P_\theta\{s_m < b\sqrt{m}\} - P_\theta\{T_+ < m, s_m < b\sqrt{m}\}. \quad (17)$$

The first terms on the right hand sides of (15)-(17) are obtained exactly in terms of the normal distribution, and the second terms may be evaluated approximately using Proposition 1. For $n \leq m$ scaled for asymptotic purposes by the relation

$$n = mr \quad \text{for some} \quad 0 < r \leq 1 \quad , \quad (18)$$

approximations to $P_\theta\{T \le n,\ s_T > 0\}$, etc. may be obtained from similar formulae but with θ_1 replaced by $\theta_1\sqrt{r}$. The approximation given by (17) and Proposition 1 is essentially that suggested in Siegmund (1977) for the type II error probability of the test defined in Section 1. It was shown to be quite accurate for computational purposes by comparing it with exact numerical results of McPherson and Armitage (1971)--see also Armitage (1975). The author has performed a small simulation to check the accuracy of the approximation suggested by (15) and Proposition 1. Although this approximation does not seem to be quite as good as that given by (17), its accuracy to within about 4-6% seems adequate for practical purposes.

Table 2 gives some examples of approximate 90% confidence intervals for θ computed from hypothetical data. The test parameters are $b = 3.45$ and $m = 148$, which give a significance level $\alpha = .01$ and power $1-\beta = .95$ at $\theta = .4$ (Armitage, 1975, p. 105). The data were selected to correspond roughly to "typical" outcomes for $\theta = .8, .6, .4, .27$, and $.2$. For example, the value $T = 19$ in the first row is the nearest integer to $E_\theta T$ for $\theta = .8$. (According to simulations and the asymptotic approximations of Siegmund, 1977, $E_\theta T \cong 18.8$ for $\theta = .8$.) The columns headed "Probability" give approximately the probability of obtaining a more extreme value than the data in the first column for the value θ equal to the one-sided confidence limit in the preceding column. Thus the first row of the third column gives approximately $P_{.35}\{T \le 19,\ s_T > 0\}$, while the corresponding entry in the fifth column gives approximately $P_{1.15}\{T \ge 19\} + P_{1.15}\{T < 19,\ s_T < 0\}$.

TABLE 2

90% CONFIDENCE INTERVALS

Data	$\underline{\theta}$	Probability	$\overline{\theta}$	Probability
$T = 19,\ s_T > 0$	.35	.053	1.15	.049
$T = 32,\ s_T > 0$	.26	.051	.89	.049
$T = 68,\ s_T > 0$	.17	.048	.60	.050
$T > 148,\ s_{148} = 40$	.11	.048	.40	.046
$T > 148,\ s_{148} = 30$	.06	.047	.34	.050

In regard to the heuristic discussion concerning proportional accuracy in Section 1, it is interesting to note that if one considers the midpoint $\hat{\theta} = (\underline{\theta} + \overline{\theta})/2$ of the intervals in Table 2 as a point estimator of θ, then the ratio $(\overline{\theta} - \underline{\theta})/\hat{\theta}$ is approximately constant ($\cong 1.1$) in the first four rows. In the last row the confidence interval is essentially the same as the standard fixed sample size interval based on the same value of s_{148}.

There is an apparent discrepancy between the confidence intervals given in Table 2 and the test of hypothesis defined in Section 1, in the sense that for the data in the last two rows of Table 2 the interval $(\underline{\theta}, \overline{\theta})$ does not contain 0, although the test of Section 1 fails to reject H_0: $\theta = 0$. Of course, the reason is that Table 2 gives 90% confidence intervals whereas the significance level of the test is $\alpha = .01$. These choices were made to provide an excuse for the following discussion.

For $2\gamma > P_0\{T \leq m\}$ a test which rejects H_0: $\theta = 0$ if and only if the $(1 - 2\gamma)$ 100% confidence interval defined above fails to contain 0 has as its rejection region $\{T \leq m\} \cup \{T > m, |s_m| \geq \xi\}$, where $0 < \xi < b\sqrt{m}$ is chosen so that

$$2\gamma = P_0\{T \leq m\} + P_0\{T > m, |s_m| \geq \xi\} \quad .$$

Along with its larger type I error probability this test has a smaller type II error probability than the test of Section 1. It has some additional flexibility which suggests that it should be studied in its own right. Since this test is defined by three parameters b, m, and ξ, it is possible to impose another constraint in addition to the usual determination of α and β. For example, b may be increased to provide for more accurate estimation of θ. Another possibility is to choose a smaller value of m than the test of Section 1 requires, so that the maximum sample size of the sequential test is not so much larger than that of the fixed sample size test of the same α and β. Both of these modifications tend to increase the expected sample size for large $|\theta|$ and must be evaluated accordingly.

The methods of this section may also be used to obtain point estimators of θ. The obvious suggestion is the midpoint $(\underline{\theta} + \overline{\theta})/2$ of the confidence interval $[\underline{\theta}, \overline{\theta}]$ for some confidence coefficient $(1-2\gamma)$ 100%. The midpoint of the 0% interval is perhaps the most natural. For the data of Table 2 the midpoints of the 0% intervals are essentially the same as the midpoints of the 90% intervals. It is not known whether this is true generally.

3. Estimators Based on $\bar{x}_T$,

The estimators of the preceding section are unusual in the sense that when $T \leq m$ they are defined directly in terms of T and only indirectly in terms of $\bar{x}_T = s_T/T$. This has important implications when σ is unknown, for then analogous arguments give estimators for the parameter $\theta = \mu/\sigma$ and not μ alone--see Section 4. In this section simple estimators based on $\bar{x}_T$, are suggested and some of their properties studied, largely by simulation.

The naive estimator $\bar{x}_T$, is rather badly biased when $|\theta|$ is large, but this bias is easily approximated and may be reduced considerably by a simple modification. To compute the asymptotic bias of $\bar{x}_T$, as $b \to \infty$ and $m \to \infty$ so that (8) holds, observe that

$$s_T/T = b/\sqrt{T} + (s_T - b\sqrt{T})/T = b/\sqrt{T} + (s_T^2/T - b^2)/(s_T + b\sqrt{T}) \quad . \tag{19}$$

From (9) it follows that

$$s_T + b\sqrt{T} \sim 2b^2/\theta \quad \text{a.s.} \quad (\theta \neq 0) \quad . \tag{20}$$

A Taylor series expansion gives

$$b/\sqrt{T} = \theta - \theta^3(T - b^2/\theta^2)/2b^2 + 3\theta^5(T - b^2/\theta^2)^2/8b^4 + \ldots \quad . \tag{21}$$

The following asymptotic results have been obtained by Lai and Siegmund (1977, unpublished) and Woodroofe (1976a) in terms of $S_n = \theta s_n - n\theta^2/2$ and $\tau_+ = \inf\{n: S_n > 0\}$:

$$E_\theta T = (b^2 - 1)/\theta^2 + E_\theta S^2_{\tau_+}/\theta^2 E_\theta S_{\tau_+} + o(1), \; \mathrm{Var}_\theta \; T \sim 4b^2/\theta^4 \quad ,$$

and $E_\theta(s_T^2/T - b^2) \to E_\theta S^2_{\tau_+}/2\, E_\theta S_{\tau_+}$ as $b \to \infty$. Substituting these results into (19) and (21) and appealing to (20) yields for $\theta \neq 0$

$$E_\theta(s_T/T) = \theta(1 + 2/b^2) + o(b^2) \qquad (b \to \infty) \quad . \tag{22}$$

It follows heuristically from (10) and (22) and rigorously after some additional calculation that

$$\begin{aligned} E_\theta \bar{x}_{T'} &= \theta(1 + 2/b^2) + o(b^{-2}) && \text{for } |\theta| > \theta_1 \\ &= \theta + o(b^{-2}) && \text{for } |\theta| < \theta_1 \end{aligned}$$

This suggests estimating θ by

$$\begin{aligned} \hat{\theta} &= (s_T/T)/(1 + 2/b^2) && \text{if } T \leq m \\ &= s_m/m && \text{if } T > m \end{aligned} \tag{23}$$

so that $E_\theta\hat{\theta} = \theta + o(b^{-2})$ for all $|\theta| \neq \theta_1$. Obviously this estimator should be altered further to avoid the embarrassing possibility of estimating $|\theta|$ to be larger when $T > m$ and $|s_m|$ is close to $b\sqrt{m}$ than when $T = m$. It is doubtful that such a refinement would significantly alter the behavior of the estimator except perhaps for $|\theta|$ close to θ_1. In any case the author has made no effort in this direction.

A similar calculation shows that

$$E_\theta(\hat{\theta} - \theta)^2 \sim E_\theta(\bar{x}_{T'} - \theta)^2 \sim (\theta/b)^2 \qquad |\theta| \geq \theta_1$$
$$\sim 1/m \qquad |\theta| \leq \theta_1 \quad . \tag{24}$$

According to (9) and a Theorem of Anscombe (1952), as $b \to \infty$ $\sqrt{T'}(\bar{x}_{T'} - \theta)$ has asymptotically a standard normal distribution. Since $b(\hat{\theta} - \bar{x}_{T'}) \to 0$ in probability, the same is true for $\sqrt{T'}(\hat{\theta} - \theta)$. Hence with z_γ defined by $1 - \Phi(z_\gamma) = \gamma$ the interval

$$[\hat{\theta} - z_\gamma/\sqrt{T'} \ , \ \hat{\theta} + z_\gamma/\sqrt{T'}] \tag{25}$$

is an approximate $(1-2\gamma)$ 100% confidence interval for θ. An alternative interval, which is slightly longer when $T \leq m$, and which emphasizes the manner in which T provides for estimating θ with prescribed proportional accuracy is

$$[\hat{\theta} - z_\gamma \max(|\bar{x}_{T'}|/b, 1/\sqrt{m}), \hat{\theta} + z_\gamma \max(|\bar{x}_{T'}|/b, 1/\sqrt{m})] \quad . \tag{26}$$

For the current problem of estimating a normal mean with known variance, the difference between (25) and (26) is practically negligible. For other parametric families and for the case of a normal mean with unknown variance, the analogous difference may be important--see Section 4.

Table 3 contains the results of a small Monte Carlo experiment to check the accuracy of the preceding asymptotic analysis. The values chosen for b and m were again 3.45 and 148 respectively. The value of

TABLE 3

ESTIMATED PROPERTIES OF $\hat{\theta}$ OBTAINED FROM
A 200 REPETITION MONTE CARLO EXPERIMENT

θ	$\hat{E}_\theta\hat{\theta}$ ± Standard Error	$\hat{E}_\theta(\hat{\theta}-\theta)^2$	Coverage Percentage
.80	.79 ± .02	.075	.895
.60	.64 ± .02	.045	.905
.40	.43 ± .01	.051	.900
.27	.30 ± .01	.029	.900
.20	.21 ± .01		.915

$1 + 2/b^2$ is 1.168, so the recommended bias reducing factor decreases s_T/T by about 14%. The value of θ_1 is .284. The row corresponding to .27 is of particular interest, for in a neighborhood of $\theta = \theta_1$ the preceding asymptotic analysis can be expected to yield good approximations only for b quite large. (Actually for $\theta = \theta_1$ the analysis given for $E_\theta(s_{T'}/T')$ breaks down. Alternative calculations yield a bias of order b^{-2}, but the details have been omitted.)

The figures in Table 1 suggest that $\hat{\theta}$ has a small positive bias, but that the bias is much smaller than that of the unmodified estimator $s_{T'}/T'$. The mean square error is poorly predicted by the asymptotic theory. However, the mean square error of the unmodified estimator $s_{T'}/T'$ (which is not reported here) seems to be consistently about 50% larger than for $\hat{\theta}$. The percentage of times the interval (26) covered the true parameter θ is remarkably close to the nominal .90.

The following discussion explores the relation between the approximate confidence intervals given by (26) and those of the preceding section. As a first step, it is interesting to compute the intervals given by (26) for the hypothetical data in Table 2. In order to do this for the data of the first three rows, it is necessary to hypothesize a value for s_T in addition to T. But $s_T = b\sqrt{T} + (s_T - b\sqrt{T})$; and it follows from results of Lai and Siegmund (1977, unpublished) or Woodroofe (1976a) that as $b \to \infty$ $E_\theta(s_T - b\sqrt{T}) \to E_\theta S_{\tau_+}^2/2\theta E_\theta S_{\tau_+}$, where $S_n = \theta s_n - n\theta^2/2$ and $\tau_+ = \inf\{n: S_n > 0\}$. It may be shown that $E_\theta S_{\tau_+}^2/2\theta E_\theta S_{\tau_+} = .584 \ldots + \theta/8 + o(\theta)$ as $\theta \to 0$, which suggests using

$$\hat{s}_T = b\sqrt{T} + .584 \tag{27}$$

as a hypothetical value for s_T as a function of T. Table 4 gives approximate 90% confidence intervals for θ for the hypothetical data of Table 1 with s_T approximated by (27). The intervals in Table 4 are

TABLE 4

APPROXIMATE 90% CONFIDENCE INTERVALS FROM (26)

Data	$\hat{\theta}$	Lower Limit	Upper Limit
$T = 19$, $s_T = 15.62$	.704	.31	1.10
$T = 32$, $s_T = 20.1$	.538	.24	.84
$T = 68$, $s_T = 29.0$	.366	.16	.57
$T > 148$, $s_{148} = 40$	.270	.13	.41
$T > 148$, $s_{148} = 30$	.203	.07	.34

about the same length as those in Table 2. For data $T \leq m$ they tend to be shifted slightly toward smaller values of $|\theta|$, which is a reflection of the fairly substantial bias reducing factor which goes into $\hat{\theta}$.

It is also possible to give a crude analytic approximation to the coverage probability of (26). It is easy to see that

$$P_\theta\{\hat{\theta} - \theta > z_\gamma |\bar{x}_T|/b,\ s_T > 0\} = P_\theta\{T < \underline{m}((s_T - b\sqrt{T})/\sqrt{T}),\ s_T > 0\} \quad , \qquad (28)$$

where $\underline{m}(x) = [\theta^{-1}\{b(1+2/b^2)^{-1} - z_\gamma\}\{1+x/b\}]^2$. If the argument of the function $\underline{m}$ on the right hand side of (28) were not random, the methods of the preceding section would give an approximation to (28). Since T is an integer valued random variable and very small changes in x do not change the integer part of $\underline{m}(x)$, it does not seem completely outrageous to replace the random variable $(s_T - b\sqrt{T})/\sqrt{T}$ by an approximation to its expectation. Equation (9) and the discussion preceding (27) suggest considering

$$P_\theta\{T < \underline{m}(b^{-1}\theta(.584 + \theta/8))\} \qquad (29)$$

as an approximation to (28). For $b = 3.45$, $m = 148$, $z_\gamma = 1.645$ and $\theta = .6$ this argument yields the approximation

$$P_\theta\{\hat{\theta} - \theta > z_\gamma |\bar{x}_T|/b\} \cong .030 \quad , \qquad (30)$$

and a similar calculation gives

$$P_\theta\{\hat{\theta} - \theta < - z_\gamma |\bar{x}_T|/b\} \cong .073 \quad , \qquad (31)$$

so that the probability that interval (26) fails to cover θ is about $.030 + .073 = .103$. The corresponding calculations for $\theta = .4$ yield respectively .034, .076, and .109. Although these calculations should not be taken too seriously, it is interesting to note that the right tail probability in (30) is smaller than the left tail probability in (31), which is consistent with the earlier observation that the intervals (26) are shifted towards smaller values of $|\theta|$ than those of Section 2. Also the total coverage probability seems to remain reasonably close to the nominal .90, which is consistent with the Monte Carlo results presented in Table 3. The author has performed a few calculations for other values of b and θ and obtained results which are reasonably consistent with those reported here.

4. Remarks on Other Parametric Models

The results of the preceding sections should extend in a fairly straightforward manner to problems involving other one parameter exponential models. At least there seem to be no conceptual difficulties, although the technical requirements of the approximate calculations of Section 2 may be considerable. The important case of matched pairs of Bernoulli outcomes may be reduced to a one parameter model by the customary practice of discarding success-sucess and failure-failure pairs (Wald, 1947, p. 107). However, numerical computation of the constants which enter into the asymptotic formulas of Section 2 may not be appreciably simpler than exact numerical computation of the probabilities from the difference equations they satisfy.

By way of contrast the case of a normal population with unknown mean and variance presents a new conceptual problem. Now assume that $x_1, x_2, \ldots$ are independent and normally distributed with unknown mean μ

and variance σ^2. Siegmund (1977) suggests a stopping rule analogous to (1) for testing H_0: $\mu = 0$ against H_1: $\mu \neq 0$ defined in terms of the log generalized likelihood ratio statistic

$$Z_n = \frac{1}{2} n \log(1 + \bar{x}_n^2 / v_n^2) \quad .$$

Here $\bar{x}_n = n^{-1} \Sigma_1^n x_k$ and $v_n^2 = n^{-1} \Sigma_1^n (x_k - \bar{x}_n)^2$. For $a > 0$ and $m_0 = 2,3,\ldots$ let $\hat{T}$ = first $n \geq m_0$ such that $Z_n > a$. For $m = m_0, m_0 + 1, \ldots$ stop sampling at $\hat{T}' = \min(\hat{T}, m)$ and reject H_0 if and only if $\hat{T} \leq m$. An asymptotic approximation for the significance level of this test as $m_0 \to \infty$, $m \to \infty$, and $a \to \infty$ in such a way that $\theta_0 = \sqrt{2a/m_0}$ and $\theta_1 = \sqrt{2a/m}$ remain fixed is given by Siegmund (1977), and an approximation to the power may also be obtained.

An attempt to adapt the estimation methods of Sections 2 and 3 leads to several new problems. The distribution of the Z_n and hence of $\hat{T}$ depends on the parameters (μ, σ) only through the value of $\theta = \mu/\sigma$, and thus the method of Section 2 yields confidence intervals for θ but not for μ alone. In a clinical trials context the primary goal of experimentation is decision oriented: to recommend the superior treatment if there is an appreciable difference in their effects; and a confidence interval provides a more useful way of measuring this difference than a simple test of hypothesis. However, the exact parameter used to measure this difference seems not to be terribly important, and for this reason a confidence interval for θ may be as useful as one for the customary μ. The computations required to give confidence intervals for θ by the method of Section 2 are quite complicated and will be considered in a future publication.

Estimation of θ based on $\bar{x}_{\hat{T}'}/v_{\hat{T}'}$ may be carried out in much the same fashion as in Section 3.

In principle μ may be estimated by $\bar{x}_{\hat{T}'}$, perhaps adjusted for bias, and approximate confidence intervals analogous to (25) or (26) may be given. However, it is not true in general that small values of $\hat{T}$ are evidence in favor of large values of $|\mu|$ (although they are evidence of large values of $|\theta|$). Two new facts make this line of attack more complicated than in the case of known σ.

Computation of the asymptotic bias of $\bar{x}_{\hat{T}}$ seems much more difficult than in the case of known σ. Presumably the bias is smaller than before because small values of $\hat{T}$ may be caused by large values of $\bar{x}_n$ or small values of v_n^2 or both. A small Monte Carlo study not reported in detail suggests that some adjustment is advisable but that the adjustment used in the case of known σ is slightly too large.

Given a satisfactory point estimator $\hat{\mu}$ for μ, analogous confidence intervals corresponding to (25) and (26) are respectively

$$[\hat{\mu} - z_\gamma v_{\hat{T}'}/\sqrt{\hat{T}'} \ , \ \hat{\mu} + z_\gamma v_{\hat{T}'}/\sqrt{\hat{T}'}] \tag{32}$$

and

$$[\hat{\mu} - z_\gamma \max(|\bar{x}_{\hat{T}'}|/\sqrt{2a} \ , \ v_m/\sqrt{m}), \ \hat{\mu} + z_\gamma \max(|\bar{x}_{\hat{T}'}|/\sqrt{2a} \ , \ v_m/\sqrt{m})] \ . \tag{33}$$

The second interval has the interpretation of estimating μ with prescribed proportional accuracy whenever $\hat{T} \leq m$. It is longer than the first interval, and perhaps considerably longer, by virtue of the inequalities

$$v_{\hat{T}}/\sqrt{\hat{T}} \leq v_{\hat{T}} \sqrt{\log\left(1 + \bar{x}_{\hat{T}}^2/v_{\hat{T}}^2\right)/2a} \leq |\bar{x}_{\hat{T}}|/\sqrt{2a} \ .$$

Determination of a reasonable estimator $\hat{\mu}$ and of the relative merits of (33) and (34) awaits further study.

Appendix

Proof That $[\underline{\theta}, \overline{\theta}]$ Has Required Coverage Probability

The notation and definitions are those of Section 2. It suffices to show for all θ that $P_\theta\{\underline{\theta} > \theta\} \leq \gamma$. That $P_\theta\{\overline{\theta} < \theta\} \leq \gamma$ is proved similarly and these two statements together imply that $P_\theta\{\underline{\theta} \leq \theta \leq \overline{\theta}\} \geq 1 - 2\gamma$. There are three cases to consider: (I) $\underline{\theta}_1(m) \leq \theta$, (II) $\underline{\theta}_1(m) > \theta$ and $\underline{\theta}_2(-b\sqrt{m}) \leq \theta$, (III) $\underline{\theta}_2(-b\sqrt{m}) > \theta$ (and necessarily $\underline{\theta}_1(m) > \theta$). In case I $P_\theta\{T \leq m,\ s_T > 0\} \geq \gamma$ and hence

$$\begin{aligned} P_\theta\{\underline{\theta} > \theta\} &= P_\theta\{T \leq m,\ s_T > 0,\ \underline{\theta}_1(T) > \theta\} \\ &= P_\theta\{T \leq n,\ s_T > 0\}, \text{ where } \underline{\theta}_1(n) > \theta \geq \underline{\theta}_1(n+1) \quad , \\ &< P_{\underline{\theta}_1(n)}\{T \leq n,\ s_T > 0\} \\ &= \gamma \quad . \end{aligned}$$

In case II $P_\theta\{T \leq m,\ s_T > 0\} < \gamma \leq P_\theta\{T \leq m,\ s_T > 0\} + P_\theta\{T > m\}$ and hence

$$\begin{aligned} P_\theta\{\underline{\theta} > \theta\} &= P_\theta\{T \leq m,\ s_T > 0\} + P_\theta\{T > m,\ \underline{\theta}_2(s_m) > \theta\} \\ &= P_\theta\{T \leq m,\ s_T > 0\} + P_\theta\{T > m,\ s_m > \xi\}, \text{ where } \underline{\theta}_2(\xi) = \theta \quad , \\ &= \gamma \quad . \end{aligned}$$

Case III is treated similarly and is omitted.

Proof of Proposition 1.

The following argument gives an asymptotic upper bound for the probability on the left hand side of (12). The corresponding lower bound is obtained by a similar but considerably easier argument. For

$k = 0,1,\ldots,m$ and $0 < x < \infty$, since s_m is sufficient for θ and (8) holds

$$P_\theta\{T_+ < m,\ b\sqrt{m} - x \le s_m < b\sqrt{m}\} \le P_\theta\{T_+ < m-k,\ b\sqrt{m} - x \le s_m < b\sqrt{m}\}$$

$$+ m^{-\frac{1}{2}}\phi(\sqrt{m}(\theta_1 - \theta))\int_0^x P_0\{s_n > b\sqrt{n} \text{ for some } m-k \le n < m \mid s_m = \theta_1 m - y\} \tag{34}$$

$$\cdot \exp\{(\theta_1 - \theta)y - y^2/2m\}dy \quad .$$

Lemma 1. For fixed $k = 0,1,\ldots,$ and $0 < y < \infty$

$$P_0\{s_n > b\sqrt{n} \text{ for some } m-k \le n < m \mid s_m = b\sqrt{m} - y\} \to P_{-\frac{1}{2}\theta_1}\{\tau(y) \le k\}$$

Proof. The conditional probability above may be rewritten

$$P_0\{s_m - s_n < b\sqrt{m} - b\sqrt{n} - y \text{ for some } m-k \le n < m \mid s_m = b\sqrt{m} - y\}$$

$$= P_0\{s_i < b\sqrt{m} - b\sqrt{m-i} - y \text{ for some } 0 < i \le k \mid s_m = \theta_1 m - y\} \quad .$$

It is easily verified by straightforward calculation that $b\sqrt{m} - b\sqrt{m-i} = \frac{1}{2}m^{-\frac{1}{2}}bi + O(m^{-3/2}b) \to \frac{1}{2}\theta_1 i$ and the joint density of $x_1,\ldots,x_i$ given $s_m = m\theta_1 - y$ converges to that of independent normal random variables with mean θ_1 and variance 1. The lemma follows easily

Lemma 2. For $0 < \theta < \theta_1$ and fixed $k = 0,1,\ldots,$

$$P_\theta\{T_+ < m-k\} \le \varepsilon_\theta(k)\ m^{-\frac{1}{2}}\ \phi((\theta_1 - \theta)\sqrt{m}) \quad ,$$

where $\varepsilon_\theta(k)$ does not depend on m and converges to 0 as $k \to \infty$.

Proof.

$$P_\theta\{T_+ < m-k\} \le \sum_{n<m-k} P_\theta\{s_n > b\sqrt{n}\} = \sum_{i=k+1}^{m-1} [1 - \Phi\{(\theta_1 - \theta)\sqrt{m} + \theta(\sqrt{m} - \sqrt{m-i})\}] \quad .$$

The inequalities $\sqrt{m} - \sqrt{m-i} \geq \frac{1}{2} m^{-\frac{1}{2}} i$ and $1 - \Phi(x) \leq x^{-1} \phi(x) \quad (x > 0)$ yield the required result with

$$\varepsilon_\theta(k) = (\theta_1 - \theta)^{-1} \sum_{i=k+1}^{\infty} \exp\{-\tfrac{1}{2}\theta(\theta_1 - \theta)i\} \quad .$$

Lemma 3. For each θ, $k = 0,1,\ldots,$ and $0 < x < \infty$

$$P_\theta\{T_+ < m-k,\ b\sqrt{m} - x \leq s_m < b\sqrt{m}\} \leq \varepsilon_\theta(k)\ \exp(x^2/2m) m^{-\frac{1}{2}}\ \phi(\sqrt{m}(\theta_1 - \theta)) \quad ,$$

where $\varepsilon_\theta(k)$ does not depend on m and converges to 0 as $k \to \infty$. For $\theta > 0$, $\varepsilon_\theta(k)$ may be chosen to be independent of x.

Proof. For $0 < \theta < \theta_1$ this result is already implied by Lemma 2. By (8) and sufficiency of s_m

$$\begin{aligned} &P_\theta\{T_+ < m-k,\ b\sqrt{m} - x \leq s_m < b\sqrt{m}\} \\ &\quad = m^{-\frac{1}{2}} \phi(\sqrt{m}(\theta_1 - \theta)) \int_0^x P_0\{T_+ < m-k \mid s_m = b\sqrt{m} - y\} \exp\{(\theta_1 - \theta)y - y^2/2m\} dy \ . \end{aligned} \tag{35}$$

Setting $\theta = \gamma < \theta_1$ in (35) gives

$$P_\gamma\{T_+ < m-k\} \geq m^{-\frac{1}{2}}\ \phi(\sqrt{m}(\theta_1 - \gamma)) \exp(-x^2/2m) \int_0^x P_0\{T_+ < m-k \mid s_m = b\sqrt{m} - y\} dy$$

and hence by Lemma 2

$$\int_0^x P_0\{T_+ < m-k \mid s_m = b\sqrt{m} - y\} dy \leq \exp(x^2/2m) \inf_{0<\gamma<\theta_1} \varepsilon_\gamma(k) \quad .$$

Substituting this inequality back into (35) gives the desired result.

That the right hand side of (12) is asymptotically an upper bound for the left hand side follows at once from (34), Lemma 1, and

Lemma 3 in the case $x < \infty$. For the case $x = \infty$, for arbitrary finite x'

$$P_\theta\{T_+ < m,\ s_m < b\sqrt{m}\} = P_\theta\{T_+ < m,\ b\sqrt{m} - x' < s_m < b\sqrt{m}\}$$
$$+ P_\theta\{T_+ < m,\ s_m < b\sqrt{m} - x'\} \quad . \tag{36}$$

For $\theta > \theta_1$

$$P_\theta\{T_+ < m,\ s_m < b\sqrt{m} - x'\} \le P_\theta\{s_m < m\theta_1 - x'\} = \Phi\{\sqrt{m}(\theta_1 - \theta) - x'/\sqrt{m}\}$$
$$\le \exp\{(\theta_1 - \theta)x'\}\ \phi(\sqrt{m}(\theta_1 - \theta))/\sqrt{m}(\theta - \theta_1) \quad ,$$

so the desired result follows from the case of finite x by first letting $m \to \infty$ and then $x' \to \infty$ in (36). The required estimate for the second probability on the right hand side of (36) in the case $0 < \theta < \theta_1$ is provided by Lemma 2 and the following result.

Lemma 4. For $0 < \theta < \theta_1$, $k = 0,1,\ldots,$ and $0 < x < \infty$ for all m sufficiently large

$$P_\theta\{m - k \le T_+ < m,\ s_m < b\sqrt{m} - x\} \le \delta(x)\ \phi(\sqrt{m}(\theta_1 - \theta))/\sqrt{m} \quad ,$$

where $\delta(x) \to 0$ as $x \to \infty$.

Proof. The proof follows easily from the argument of Lemma 2 and the inequality

$$P_\theta\{m - k \le T_+ < m,\ s_m < b\sqrt{m} - x\} \le \sum_{m-k\le n<m} P_\theta\{T_+ = n\}P_\theta\{s_m - s_n < b\sqrt{m} - b\sqrt{n} - x\}$$
$$\le \sum_{0<i\le k} \{1 - \Phi(b - \theta\sqrt{m-i})\}\Phi((\theta_1 - \theta)\sqrt{i} - x/\sqrt{i})$$

References

Anscombe, F.J. (1952). Large sample theory of sequential estimation, Proc. Cambridge Philos. Soc. **48**, 600-607.

Anscombe, F.J. (1953). Sequential estimation, J. Roy. Statist. Soc. Ser. B **15**, 1-21.

Armitage, P. (1975). Sequential Medical Trials, 2nd ed., Oxford: Blackwell.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II, John Wiley and Sons, New York.

McPherson, C.K. and Armitage, P. (1971). Repeated significance tests on accumulating data when the null hypothesis is not true, J. Roy. Statist. Soc. Ser. A **134**, 15-26.

Siegmund, D. (1975). Error probabilities and average sample number of the sequential probability ratio test, J. Roy. Statist. Soc. Ser. B **37**, 394-401.

Siegmund, D. (1977). Repeated significance tests for a normal mean, Biometrika **64**, 177-189.

Wald, A. (1947). Sequential Analysis, John Wiley and Sons, New York.

Woodroofe, M. (1976a). A renewal theorem for curved boundaries and moments of first passage times, Ann. Prob. **4**, 67-80.

Woodroofe, M. (1976b). Frequentist properties of Bayesian sequential tests, Biometrika **63**, 101-110.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (14) No. 2 TR-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) ESTIMATION FOLLOWING SEQUENTIAL TESTS.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) D./Siegmund		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-77-C-0306, PHS-GM-21895-08
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ~~(NR-042-373)~~ (11) 1 Oct 77
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. ~~REPORT DATE~~ October 1, 1977
		13. NUMBER OF PAGES 26 (12) 30p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release and sale; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

SEQUENTIAL TEST

CONFIDENCE INTERVAL

STOPPING RULE

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two methods for obtaining sequential test based confidence intervals for the mean of a normal distribution are proposed and compared. One method involves reducing the bias introduced into the sample mean by optional stopping. The other defines the confidence interval directly in terms of the stopping time of the test.

332 580 Gm